

Sozialwissenschaften

Quandt, Markus; Mauer, Reiner

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Quandt, M., & Mauer, R. (2012). Sozialwissenschaften. In H. Neuroth, S. Strathmann, A. Oßwald, R. Scheffel, J. Klump, & J. Ludwig (Hrsg.), *Langzeitarchivierung von Forschungsdaten: eine Bestandsaufnahme* (S. 61-81). Boizenburg: Hülsbusch. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-457735>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-ND Lizenz (Namensnennung-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-ND Licence (Attribution-NoDerivatives). For more Information see: <https://creativecommons.org/licenses/by-nd/4.0>

4 Sozialwissenschaften

Markus Quandt, Reiner Mauer

4.1 Einführung in den Forschungsbereich

Der Begriff der Sozialwissenschaften bezeichnet keine in sich geschlossene Wissenschaftsdisziplin, sondern ist ein Sammelbegriff für viele akademische Fächer von der Anthropologie bis zur Volkswirtschaftslehre.¹ Als Gemeinsamkeit dieser Fächer kann man sehen, dass sie Fragen des gesellschaftlichen Zusammenlebens behandeln – etwa solche der sozialen Gerechtigkeit, der Analyse politischer Systeme, der Erklärung sozialer Probleme wie lokal erhöhter Kriminalitätsraten, oder auch der Prozesse in kleinen Gruppen. Entsprechend sind die Kernfächer der Sozialwissenschaften die Soziologie und die Politologie, die selbst wiederum zahlreiche Subdisziplinen haben.

Die für das Thema *Forschungsdaten* relevante Achse zur näheren Bestimmung des Faches ist jedoch sicher die der typischen Datenverwendung. Dies führt auf den Begriff der *empirischen Sozialforschung* als Untermenge der Sozialwissenschaften, in der eine gewisse Einheitlichkeit der Wissenschaftsauffassung und der anzuwendenden Methoden besteht, aber zugleich immer noch eine enorme thematische Offenheit. Dementsprechend behandelt empirische Sozialforschung etwa Bereiche wie die Bildungsforschung (geteilt mit der Disziplin der Erziehungswissenschaften, s. Kap. 6 in diesem Band) oder die Sozialpsychologie (geteilt mit der Disziplin der Psychologie). Eine wichtige Eigenheit des Untersuchungsgegenstandes der sozialen Phänomene ist es dabei, dass sozialwissenschaftliche Forschungsdaten sehr oft auch (partielle) Momentaufnahmen

1 Im englischen Sprachraum bezeichnen Social Sciences noch weit stärker als im deutschsprachigen Raum auch die Volkswirtschaftslehre. Auch die DFG führt die Wirtschaftswissenschaften in der Gruppe „Geistes- und Sozialwissenschaften“.

bestimmter gesellschaftlicher Zustände sind. Diese können im Rückblick ohnehin schon neben ihrem Wert für die ursprüngliche Forschungsfrage sehr oft eine historisch beschreibende Funktion bekommen. Durch die systematische Einordnung in Zeitvergleiche oder -reihen erhalten sie aber einen besonderen Zusatzwert daraus, dass nur auf ihrer Grundlage dynamische Prozesse verstanden werden können. Sowohl historische amtliche Statistiken wie auch Daten aus Umfragen oder Beobachtungen können so einen analytischen Wert erhalten, der zum Zeitpunkt ihrer Erhebung noch nicht vorauszusehen war.

Die Projektstrukturen sind in den Sozialwissenschaften mit Ausnahme weniger Leuchtturmprojekte eher kleinteilig. So war z.B. einerseits die gesamte Fördersumme durch die DFG mit 129,7 Mio. EUR im Jahr 2008 für die Sozial- und Verhaltenswissenschaften (einschließlich Wirtschaftswissenschaften) im Vergleich zu den meisten anderen Fachgebieten recht gering, zum anderen ist die Relation von DFG-Mitteln zu Mitarbeiterzahlen um ca. ein Drittel niedriger als im Durchschnitt der DFG-Projektförderung, sozialwissenschaftliche Projekte sind also auch im Durchschnitt wenig mittelaufwändig.² Dabei sind die Sozial- und Politikwissenschaften (in der Definition des Statistischen Bundesamtes) mit über 3.300 Professoren und wissenschaftlichen Mitarbeitern zuzüglich 12.500 Wirtschaftswissenschaftlern an deutschen Hochschulen im Jahr 2009 durchaus ein mittelgroßes Fach.³ Auch außerhalb der Universitäten sind die Strukturen in Deutschland dergestalt, dass Forschungsinstitute bereits mit wenigen Dutzend Mitarbeiterinnen und Mitarbeitern als groß gelten können.

4.2 Kooperative Strukturen

Typisch für die empirische Sozialforschung sind einerseits die oben genannten kleinteiligen Projektstrukturen, in denen Analysen und Publikationen von Einzelautoren oder sehr kleinen Forschergruppen erzeugt werden. Soweit nicht Daten aus amtlichen Quellen oder prozessproduzierte Daten verwendet werden (s. den folgenden Abschnitt) ist zudem

2 Vgl. Statistisches Bundesamt (2010b), S. 164.

3 Vgl. Statistisches Bundesamt (2010a), S. 25.

die eigentliche Datenerhebung durch Befragungen oft an Meinungs- und Marktforschungsfirmen als kommerzielle Dienstleister ausgelagert. Andererseits ist hervorzuheben, dass in Publikationen häufig Daten verwendet werden, die entweder nicht eigens zur Beantwortung der jeweiligen Forschungsfrage erhoben wurden oder die aus Datenerhebungen stammen, die von vornherein zur Beantwortung mehrerer, mitunter unverbundener, Forschungsfragen durchgeführt werden. Im ersten Fall handelt es sich beispielsweise um Daten aus der originär zu Verwaltungs- oder Planungszwecken erhobenen amtlichen Statistik, im zweiten Fall um Mehrthemenumfragen. Gerade mit den letzteren stellt die Sozialforschung seit mehreren Jahrzehnten z.B. im Rahmen von „General Social Surveys“ (in Deutschland der ALLBUS⁴, europaweit vergleichend der European Social Survey, weltweit etwa das International Social Survey Programme oder verschiedene „Barometer“-Studien⁵) gezielt Kollektivgüter für spätere Analysen durch die ganze wissenschaftliche Community zur Verfügung. In jüngerer Zeit werden mit Blick auf eine vorrangige Sekundär-Nutzung zunehmend methodisch höchst aufwändige Erhebungsdatensätze etwa das Sozioökonomische Panel (SOEP)⁶, der Survey on Health, Ageing and Retirement in Europe (SHARE) oder pairfam⁷ etabliert. Derartige Projekte sind von kleinen Teams nicht mehr zu bewältigen und erfordern relativ feste organisatorische Strukturen.

Diese Orientierung auf die Sekundäranalyse von Daten durch Forscher, die nicht selbst an der Erhebung beteiligt sind, ist besonders in international vergleichenden Projekten seit langem stark ausgeprägt. Eine international vergleichende Perspektive hat sich in der empirischen Sozialforschung seit den 1960er-Jahren fast schon als Regelfall durchgesetzt, weil sich das beste Verständnis der Wirkungsweise etwa von gesellschaftlichen Institutionen, ökonomischen Ungleichheitsstrukturen oder Wertesystemen aus Vergleichen gewinnen lässt. Die Notwendigkeit von Vergleichen bedeutet sehr oft, dass Daten aus verschiedenen Gesellschaften vorliegen müssen. Damit hat eine dominante Forschungsperspektive auch die Herausbildung wenigstens informeller internationaler Datenin-

4 Vgl. Koch; Wasmer (2004).

5 Vgl. Mochmann (2008).

6 Vgl. Wagner; Frick; Schupp (2007).

7 Vgl. Huinink et al. (2010).

frastrukturen frühzeitig angetrieben, indem es infolge von Kooperationen auf der Datenerhebungsseite auch solche in der Datenaufbereitung und -bereitstellung gab und gibt, die institutionell jedoch hauptsächlich von national verankerten Instituten wie GESIS in Deutschland oder dem Interuniversity Consortium for Political and Social Research in Michigan, USA, und etlichen weiteren getragen werden.⁸

Auf nationaler Ebene wurden hingegen bereits kurz nach dem 2. Weltkrieg (Roper Center, USA) sozialwissenschaftliche Datenarchive als Dienstleister speziell für die Umfrageforschung etabliert, in Europa war das deutsche Zentralarchiv für empirische Sozialforschung (jetzt GESIS Datenarchiv für Sozialwissenschaften) ab 1960 die erste dieser Einrichtungen.⁹ Diese Datenarchive gewährleisteten gleichermaßen die Langzeitsicherung der Daten wie die zentrale Datenbereitstellung an die nationale Forschergemeinschaft, z.T. auch mit internationaler Reichweite. Freilich sind diese Institutionen überwiegend aus bestimmten Traditionen hervorgegangen und haben zum Teil die damit gesetzten Spezialisierung – überwiegend auf die akademisch geprägte Umfrageforschung – bis heute bewahrt. Entsprechend konnten bis vor kurzem jene Subdisziplinen der Sozialwissenschaften weniger gut bedient werden, die sehr stark auch andere Datentypen wie etwa Quellen aus der amtlichen Statistik oder qualitative Daten nutzen. Ferner sind nach wie vor die direkt bei den Auftraggebern von Forschungsprojekten, darunter auch Ministerien und Stiftungen, oder die bei kommerziellen Umfrageinstituten lagernden Daten in der Regel der Forschung bestenfalls unsystematisch zugänglich. Weder sind die Gesamtbestände an derartigen „verborgenen“ Daten abschätzbar, noch können Aussagen über die Langzeitsicherheit solcher Bestände gemacht werden.

Direkter Ausweis eines auch in der Forschungspolitik gestiegenen Bewusstseins für Probleme des Datenzugangs war die 2001 vom Bundesministerium für Bildung und Forschung (BMBF) eingesetzte „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissen-

8 Vgl. Mochmann (2008).

9 Mit dem Roper Center wurde 1946 das erste sozialwissenschaftliche Datenarchiv weltweit eröffnet, 1960 folgte mit dem Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln die erste Gründung in Europa. Mitte der 1960er-Jahre folgte eine Reihe weitere Gründungen in den USA und in Europa (vgl. Scheuch (2003)).

schaft und Statistik‘. In der Folge des von der Kommission vorgelegten Gutachtens¹⁰ wurde vom BMBF 2004 erstmals der Rat für Sozial- und Wirtschaftsdaten (RatSWD) berufen, der als unabhängiges Expertengremium in erster Linie eine begutachtende und koordinierende Funktion beim Aufbau von Forschungsinfrastrukturen einnimmt. Durch die parallel in Deutschland etablierten Forschungsdatenzentren und Datenservicezentren wurde insbesondere der Zugang zu solchen Daten verbessert oder erst ermöglicht, die bei Ämtern und Behörden wie der Bundesagentur für Arbeit oder den statistischen Ämtern bzw. bei Körperschaften wie der Rentenversicherung aus der Durchführung öffentlicher Aufgaben entstehen (s. Kap. 4.3). Diese Zentren sind für sich genommen kleinere Einheiten, jeweils bei der datenerzeugenden oder -haltenden Organisation angesiedelt und haben einen thematisch engen Fokus, um eine hohe fachwissenschaftliche Qualität der Datenhandhabung zu gewährleisten. Damit wird für Deutschland die Lücke zunehmend geschlossen, die neben dem breiten, aber überwiegend auf Umfragedaten gerichteten Datenservice-Angebot des GESIS-Datenarchivs bestand.^{11,12}

Für den in den letzten Jahren stetig an Bedeutung gewinnenden Bereich der qualitativen Sozialforschung (mit Datentypen wie beispielsweise Tonaufnahmen, Videoaufnahmen oder Transkriptionen von Interviews) fehlt es dagegen noch weitgehend an Einrichtungen, die sich systematisch der Archivierung und Bereitstellung solcher Daten widmen.¹³

Insgesamt hat sich die sozialwissenschaftliche Dateninfrastruktur auf nationaler Ebene in den letzten Jahren dynamisch weiterentwickelt. Diese grundsätzlich positive Entwicklung führt allerdings auch zu einer gewissen Unübersichtlichkeit im Datenzugang und birgt beispielsweise das Potential zu Doppelungen in der Datenhaltung und den daraus resultierenden Unsicherheiten über ‚autoritative‘ Versionen. Mit der Etablierung einer Datenregistrierungsagentur für sozialwissenschaftliche Forschungs-

10 Vgl. Kommission zur Verbesserung der Informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001).

11 Vgl. Wissenschaftsrat (2009); RatSWD (2010a).

12 Langfristsicherung von sozialwissenschaftlichen Forschungsdaten ist eine allgemeine Dienstleistung des GESIS-Datenarchivs. Bei Umfragedaten, historischen Zeitreihen/Aggregatdaten und biografischen Daten bestehen jedoch Spezialisierungen in Aufbereitung und Bereitstellung.

13 Vgl. Medjedović; Witzel (2010), S. 88; Wissenschaftsrat (2011c), S. 9 und 11.

daten versucht GESIS dem im ersten Schritt entgegenzuwirken und die Voraussetzungen für eine dauerhafte Identifizierung und Lokalisierung der reichhaltigen Datenbestände zu schaffen.¹⁴

Auf internationaler Ebene haben sich bis vor kurzem nur begrenzt institutionelle Strukturen ergeben, die Kooperationen zwischen Datenserviceeinrichtungen erleichtert haben. Auf europäischer Ebene wurde 1976 das Council of European Social Science Data Archives (CESSDA¹⁵) als freiwilliges Kooperationsnetzwerk nationaler Datenserviceeinrichtungen gegründet, ein Jahr später folgte die International Federation of Data Organizations for the Social Sciences (IFDO¹⁶).¹⁷ Nach der Aufnahme von CESSDA auf die ESFRI-Roadmap¹⁸ ist CESSDA nun auf dem Weg zu einem sog. European Research Infrastructure Consortium (ERIC¹⁹) und wird damit zu einer Körperschaft europäischen Rechts, die dauerhaft zentrale Koordinationsfunktionen für die sozialwissenschaftliche Datenlandschaft in Europa übernehmen und sogar eigene technische Infrastrukturen betreiben kann.

4.3 Daten und Metadaten

Sozialwissenschaftler verwenden eine Vielzahl an unterschiedlichen Methoden um Daten zur Untersuchung sozialer Phänomene zu gewinnen. Das Spektrum reicht von großen standardisierten Umfragestudien mit repräsentativen Bevölkerungsstichproben, die in Form von einfachen Querschnittstudien aber mit zunehmender Bedeutung als Längsschnittstudien (Trendstudien, Panelstudien) oder ländervergleichenden Studien (bzw. Kombinationen dieser Designs) erhoben werden über semi- bzw. unstrukturierte Interviews und Textdaten, die mit inhaltsanalytischen Verfahren analysiert werden (wie bpsw. Parteiprogrammen) bis hin zu

14 Vgl. GESIS (2011a). Seit 2011 wird der Service in Kooperation mit der Deutschen Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft (ZBW) auch auf wirtschaftswissenschaftliche Daten ausgeweitet.

15 Vgl. CESSDA (2011c).

16 Vgl. IFDO (2011).

17 Vgl. Mochmann (2008).

18 Vgl. ESFRI (2006).

19 Vgl. European Commission (2009).

Daten, die durch Experimente und Beobachtung gewonnen werden. Darüber hinaus spielen auch Daten eine sehr wichtige Rolle, die nicht primär für Forschungszwecke erhoben werden. Dazu zählen insbesondere Daten der amtlichen Statistik und zunehmend auch prozessproduzierte Daten, die im Rahmen von Verwaltungsprozessen bspw. bei Renten-, Kranken- und Arbeitslosenversicherung entstehen. Neuere Formen von Daten wie Transaktionsdaten²⁰, die bspw. bei der Verwendung von Kreditkarten, Mobiltelefonen und insbesondere auch bei der Nutzung des Internets entstehen oder die Erhebung biologischer Merkmale (bzw. Cortisol-Level) im Rahmen von Umfragestudien²¹ erweitern das Datenportfolio empirisch arbeitender Sozialwissenschaftler.²² Dementsprechend fallen sozialwissenschaftliche Forschungsdaten in sehr unterschiedlichen Formen und Formaten an.

Eine wichtige Rolle für die empirischen Sozialwissenschaften spielen insbesondere sogenannte Mikrodaten. Diese zeichnen sich dadurch aus, dass sie Informationen zu jeder Untersuchungseinheit enthalten. Eine solche Untersuchungseinheit wird im häufigsten Fall ein Individuum, kann bspw. aber auch ein Haushalt oder ein Unternehmen sein.²³ Neben den großen wissenschaftlichen Datenerhebungsprogrammen werden wichtige Mikrodaten vor allem durch die amtliche Statistik erstellt (z.B. Mikrozensus oder Einkommens- u. Verbrauchsstichprobe) oder entstehen im Rahmen von Verwaltungsakten (z.B. Rentenversicherung).

Publikation und Langzeitarchivierung von Forschungsdaten

Wie oben dargestellt existiert neben dem GESIS Datenarchiv mittlerweile eine ganze Reihe von Forschungsdatenzentren. Bei aller Heterogenität hinsichtlich der betreuten Daten und Themen ist diesen Institutionen gemein, dass sie einen starken Fokus auf Produktion und Anreicherung

20 Vgl. hierzu z.B. Lane (2009).

21 Vgl. z.B. Schnell (2009).

22 Einen fundierten Überblick über die verschiedenen „neuen“ Datenformen und eine umfassende Darstellung der Situation der sozialwissenschaftlichen Dateninfrastruktur bietet das vom RatSWD herausgegebene zweibändige Kompendium „Building on Progress“ (vgl. RatSWD (2010a)).

23 Im Gegensatz zu Mikrodaten beinhalten Makro- oder Aggregatdaten zusammengefasste Versionen solcher Informationen in der Form von Mittelwerten, absoluten oder relativen Häufigkeiten oder anderen statistischen Maßzahlen.

qualitativ hochwertiger Daten und insbesondere auf eine Verbesserung der Zugänglichkeit solcher Daten setzen. Fragen der Langzeitarchivierung sind demgegenüber allerdings bisher nicht in systematischer Weise für alle Bereiche der Sozialwissenschaften adressiert worden. Einzig das Datenarchiv der GESIS bietet gegenwärtig ausdrücklich einen Langzeitarchivierungsservice insbesondere für – aber nicht beschränkt auf – Umfragedaten an.

Die momentan 18 beim RatSWD akkreditierten Forschungsdatenzentren und Datenservicezentren²⁴ stellen die von ihnen – z.T. selbst produzierten – aufbereiteten und dokumentierten Datenbestände i.d.R. als faktisch anonymisierte scientific use files oder auch als absolut anonymisierte public use files für wissenschaftliche Zwecke zur Verfügung und bieten Beratung und Unterstützung bei der Nutzung. Für datenschutzrechtlich sensiblere Daten stehen entweder Arbeitsplätze im jeweiligen Datenzentrum zur Verfügung oder es werden Möglichkeiten zur kontrollierten Datenfernverarbeitung geboten. Häufig muss aufgrund der besonderen Natur der in den Forschungsdatenzentren angebotenen Daten (wie bspw. im Fall der Mikrodaten der amtlichen Statistik) die Nutzung beantragt werden und setzt den Abschluss eines Nutzungsvertrages voraus.

Im Datenarchiv der GESIS werden vorrangig Umfragedaten und Daten der historischen Sozialforschung nach internationalen Archivstandards aufbereitet, dokumentiert und langfristig gesichert. Alle Daten erhalten persistente Identifikatoren (DOIs) und werden über Portale (auch mit der Möglichkeit zu einfachen Online-Analysen) oder ftp zum Download angeboten sowie auf Offline-Medien bereitgestellt. Zur Identifikation relevanter Daten werden verschiedene Recherchesysteme angeboten. So liegen zu jeder Studie detaillierte und standardisierte deskriptive und strukturelle Metadaten vor, die sich über den sogenannten Datenbestandskatalog online durchsuchen lassen. Für die wichtigsten Studien sind Recherchen bis auf die Ebene einzelner Variablen möglich.

Ausgangspunkt für die Arbeit des Archivs ist neben der Forderung nach Reproduzierbarkeit und intersubjektiver Überprüfbarkeit von empirischen Forschungsergebnissen insbesondere die Bereitstellung von Daten für neue Forschungszwecke mit neuen Auswertungszielen bzw.

24 Vgl. RatSWD (2011a).

-methoden.²⁵ Konsequenterweise steht neben der klassischen Langzeitar-
chivierung mit ihren, auf den Erhalt der Interpretierbarkeit gerichteten,
Maßnahmen vor allem das Kuratieren von Forschungsdaten im Mittel-
punkt der Arbeit. Um eine möglichst breite und optimale Nachnutzung
zu ermöglichen, werden für die Forschung besonders relevante Daten mit
teils erheblichem Aufwand mit Mehrwert versehen. Seit 2009 hat GESIS
diese Arbeiten ebenfalls überwiegend in den Rahmen von Forschungsda-
tenzentren gefasst. So werden etwa Daten der internationalen Umfrage-
forschung, Referenzstudien wie der ALLBUS oder auch Daten aus dem
Bereich der politischen Soziologie – häufig in Kooperation mit den Pri-
märforschern oder anderen Experten – in Datenkollektionen zusammen-
gefasst und besonders betreut. Variablen werden harmonisiert und stan-
dardisiert, um Vergleiche über die Zeit oder über regionale Einheiten
hinweg zu ermöglichen, und umfangreiche Metadaten und Dokumenta-
tionen zu diesen Daten erzeugt.

Mindestanforderungen für die Aufnahme von Forschungsdaten

Bisher gibt es neben dem Datenarchiv der GESIS keine sozialwissen-
schaftliche Einrichtung, die in größerem Umfang und thematischer Brei-
te Forschungsdaten archiviert. Ein Großteil der o.g. Forschungsdaten-
zentren betreut ausschließlich Daten, die sie entweder – z.B. im Rahmen
ihres gesetzlichen Auftrages – selbst erzeugen (wie bspw. die Forschungs-
datenzentren der Statistischen Ämter) oder aber Daten mit einem eng
begrenzten thematischen Fokus (wie bspw. das FDZ am Institut zur
Qualitätsentwicklung im Bildungswesen, das Daten aus den großen nati-
onalen und internationalen Schulleistungstudien, wie IGLU oder PISA,
betreut). Für derartige Datenzentren spielt die Definition von (allgemei-
nen) Mindestanforderungen für die Aufnahme von Daten keine bedeu-
tende Rolle.

Am GESIS Datenarchiv werden Forschungsdaten entweder gezielt ak-
quiriert oder aber wie in der Mehrzahl der Fälle dem Archiv von Primär-
forschern zur Archivierung angeboten bzw. die Übergabe erfolgt mehr
oder weniger routinemäßig auf der Grundlage längerfristiger Kooperatio-
nen. Grundsätzlich werden maschinenlesbare Datensätze aus allen Berei-

25 Vgl. Kaase (1999), S. 64 ff.

chen der Sozialwissenschaften archiviert, wenn a) die Studien Aussagen über die deutsche Bevölkerung oder über Teile von ihr erlaubt, b) an der Untersuchung deutsche Forscher beteiligt waren, unabhängig davon, ob sie sich auf Deutschland bezieht oder nicht und c) die Studie ganz allgemein für die sozialwissenschaftliche Forschung von Interesse sein könnte. Die gezielte Akquisition richtet sich meist auf Daten zu bestimmten Forschungsgebieten (bspw. der international vergleichenden Sozialforschung) oder besonders prominente und bedeutende Erhebungsprojekte. Neben den zuvor genannten Kriterien müssen ins Archiv aufzunehmende Studien bestimmten formalen und technischen Anforderungen entsprechen. Insbesondere müssen neben den Datensätzen selbst auch alle für eine Sekundärnutzung notwendigen Materialien vorhanden sein und an das Archiv übergeben werden (Erhebungs- bzw. Messinstrumente, Codepläne, Methodenberichte etc.). Diese den Datensatz begleitenden Materialien und Dokumente bilden auch die Grundlage für die vom Datenarchiv im Rahmen der Archivierung erzeugten standardisierten Metadaten. Die Daten selbst müssen nicht unbedingt in einem bestimmten Format vorliegen (solange sie lesbar sind), sondern werden im Zuge der Archivierung in langfristsicherungsfähige Formate überführt und für den Datenservice wiederum in von den Nutzern gewünschten Formaten bereitgestellt. Vor der endgültigen Aufnahme von Daten ins Archiv wird im Rahmen des Ingestprozesses eine Eingangskontrolle durchgeführt, die sich auf die Erhebungsinstrumente, die Dokumentation und Daten bezieht und sicherstellt, dass die Informationen vollständig und lesbar sind und mit den Daten übereinstimmen. Die Kontrolle der Daten reicht von der Überprüfung der Anzahl der Fälle, dem Abgleich der Variablen mit dem Erhebungsinstrument bis hin zu Plausibilitäts- und Konsistenzkontrollen sowie der Überprüfung hinsichtlich datenschutzrechtlicher Aspekte. Eine Übernahme der Daten in das Langzeitarchiv erfolgt erst, wenn alle im Rahmen der Eingangskontrolle festgestellten Probleme bzw. Fehler behoben werden konnten (zumindest soweit, dass eine Nutzung der Daten durch Dritte möglich und in der Folge eine Langzeitarchivierung auch sinnvoll ist). In jüngerer Zeit ist das Archiv dazu übergegangen die Aufnahme in das Langzeitarchiv durch Archivierungsverträge mit dem Datengeber auch rechtlich zu formalisieren. Dort werden insbesondere Fragen zu Verfügungs- und Nutzungsrechten sowie zum Datenschutz geklärt.

Insgesamt kann man feststellen, dass die Qualität der in das Archiv aufgenommenen Daten im Hinblick auf technische und formale Aspekte im Verlauf der letzten Dekade deutlich angestiegen ist – auch wenn es in der Breite der Projekte sicher noch deutliches Potential für Verbesserung gibt. Nach wie vor unklar ist häufig der rechtliche Status von Forschungsdaten (bspw. Wem gehören die Daten und wer darf wem, welche Nutzungsrechte einräumen.). In diesem Zusammenhang ist sehr begrüßenswert, dass Forschungsförderer – wie bspw. die DFG²⁶ – zunehmend dazu übergehen, bereits bei der Antragstellung dezidierte Aussagen zum Umgang mit Forschungsdaten zu fordern. Aus Sicht der Autoren wären allerdings konkretere und standardisierte Vorgaben erstrebenswert. Dabei ist es jedoch zunächst eine Aufgabe der einzelnen Fach-Communities und ihrer jeweiligen Datenservice-Einrichtungen, Vorschläge für die Ausgestaltung bzw. Elemente solcher Datenmanagementpläne und best practices für den Umgang mit Forschungsdaten zu erarbeiten.

Datenvolumen

Das in den Sozialwissenschaften anfallende Datenvolumen ist derzeit nicht quantifiziert. Allgemein lässt sich sagen, dass die numerischen Datensätze der empirischen Sozialforschung tendenziell eher klein sind. So summieren sich bspw. die im GESIS Datenarchiv gehaltenen Bestände – mit Daten und Dokumenten zu rd. 6.000 sozialwissenschaftlichen Studien – auf weniger als 1 TB, die sich auf knapp 600.000 Objekte verteilen. Selbst wenn man die verschiedenen on- und offline Spiegelungen bzw. Replikate des Archivs mitrechnet, ergeben sich gemessen an heutigen Speicherdimensionen keine signifikanten Größen. Allerdings lassen sich aus den genannten Zahlen nur sehr eingeschränkt Rückschlüsse auf das Gesamtvolumen an Forschungsdaten in den Sozialwissenschaften ziehen.

Datenformate

Quantitative Daten werden i.d.R. mit speziellen Statistiksoftwarepaketen verarbeitet und analysiert. In der akademischen Forschung sind dabei die Statistikpakete SPSS und Stata am weitesten verbreitet. Zunehmend ge-

26 Vgl. DFG (2010c), S. 31f.

winnt auch die Open Source Alternative ‚R‘ an Bedeutung. In der kommerziellen Forschung spielt neben den zuvor genannten Paketen auch SAS eine wichtige Rolle. Dementsprechend werden quantitative Daten in der Regel in den meist proprietären Formaten dieser Softwarepakete erstellt, analysiert und weitergegeben. Bei den am meisten genutzten Formaten handelt es sich um proprietäre Binärformate – also um nicht oder nur eingeschränkt für die Langzeitarchivierung geeignete Formate. Im Rahmen der Langzeitarchivierung ist es daher erforderlich, diese in langfristsicherungsfähige Formate umzuwandeln. Dies kann jedoch i.d.R. ohne Informationsverlust bzw. unter Erhaltung der signifikanten Eigenschaften bewältigt werden (bspw. durch Überführung der Daten in csv-Formate und der mit den Daten verbundenen internen Metadaten in textbasierte Steuerungsdateien)

Mit dem Bedeutungszuwachs der qualitativen Methoden bzw. von mixed-method Ansätzen und der generellen Tendenz zu komplexeren Studiendesigns sowie den bereits oben beschriebenen neuen Datenformen in den Sozialwissenschaften werden auch andere Formate (Audio, Video, Bildformate, Datenbanken, GIS) zunehmend wichtig und erfordern neue oder angepasste Strategien des Datenmanagements und der Langzeitarchivierung.

Nutzungsbeschränkungen

Mit Blick auf bestehende Zugangsrestriktionen zu Forschungsdaten in den Sozialwissenschaften findet sich eine ganze Bandbreite von völlig offenem und unbeschränktem Zugang für jedermann (z.B. Public Use Files) über die eingeschränkte Nutzung ausschließlich für wissenschaftliche Zwecke (z.B. Scientific Use Files) bis hin zu hochrestriktiver Nutzung (bspw. nur in extra dafür eingerichteten Räumen der datenhaltenden Institution) bzw. auch vollkommenem Ausschluss der Nutzung durch Dritte. Neben urheberrechtlich bedingten Einschränkungen handelt es sich in bestimmten Bereichen vor allem um gesetzliche und dabei insbesondere datenschutzrechtlich begründete Nutzungsrestriktionen. Dies trifft vor allem für den gesamten Bereich der prozessproduzierten Daten von Verwaltungen sowie die Mikrodaten der amtlichen Statistik zu. Immer dort, wo Daten sich auf Individuen, Betriebe o.ä. beziehen, gilt es deren Rechte zu

wahren und sie vor einer De-Anonymisierung durch Unbefugte zu schützen. So bietet bspw. das Forschungsdatenzentrum der Statistischen Ämter den Zugriff auf lediglich formal anonymisierte Daten ausschließlich auf dem Weg der kontrollierten Datenfernverarbeitung an.²⁷ Dabei kommt der Nutzer mit den Daten nicht unmittelbar in Kontakt. Vielmehr übermittelt der Nutzer lediglich seine Auswertungsprogramme und das FDZ wertet mit diesen die Originaldaten aus. Nach einer Geheimhaltungsprüfung werden die Auswertungsergebnisse wiederum dem Nutzer übergeben. Faktisch anonymisierte Daten dagegen – also Daten, bei denen Befragte nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft zu identifizieren wären – werden in Form von standardisierten Scientific Use Files auf Antrag für wissenschaftliche Zwecke auch für eine Nutzung außerhalb der geschützten Räume der Statistischen Ämter bereitgestellt. Darüber hinaus werden absolut anonymisierte Daten in Form von Public Use Files für die breite Öffentlichkeit freigegeben.

Auch bei Nutzung von Umfragedaten der akademischen Forschung sind selbstverständlich die datenschutzrechtlichen Bestimmungen zu beachten. Wenngleich dazu einschränkend anzumerken ist, dass trotz des theoretischen Risikos der Identifizierung von Befragten, die konkrete und praktische Gefahr für viele – aber keineswegs für alle! – Datensätze minimal ist bzw. gegen Null tendiert. Diese werden i.d.R. ohnehin nur in Formen weitergegeben, die als mindestens faktisch anonymisiert anzusehen sind. Viele Datensätze können sogar aufgrund ihres Designs als einfache Querschnittstudien mit einem relativ kleinen Umfang an soziodemografischen Informationen sowie ihrer relativ kleinen Stichprobengröße als absolut anonymisiert gelten. Dementsprechend kann ein Großteil der Bestände des Datenarchivs der GESIS für Zwecke der wissenschaftlichen Forschung oder Lehre direkt, d.h. ohne vorherige Bewilligung durch den Datengeber, genutzt werden. Bei ca. einem Drittel der Studien muss vor einer Nutzung eine Genehmigung des Datengebers eingeholt wird, die aber im Regelfall auch gewährt wird. Der Grund für diese Einschränkung, die ausschließlich auf Wunsch des Datengebers Anwendung findet, dürfte häufig dem Wunsch von Primärforschern entspringen, eine gewisse Kontrolle über die Nutzung der Daten zu behalten.

27 Vgl. Statistische Ämter des Bundes und der Länder – Forschungsdatenzentren (2011b).

Nachnutzung und Sekundärnutzung von Forschungsdaten

Wie oben bereits dargestellt, gibt es in den Sozialwissenschaften eine lange Tradition der Nachnutzung von Forschungsdaten. Sogenannte Sekundäranalysen, also die erneute Analyse von Daten aus bereits durchgeführten Untersuchungen mit zumeist neuer Fragestellung, gehören zum selbstverständlichen Repertoire empirisch arbeitender Sozialwissenschaftler. Dies liegt neben forschungsökonomischen Motiven – hochwertige Datenerhebungen sind mit erheblichem finanziellen und zeitlichen Aufwand verbunden – vor allen Dingen auch darin begründet, dass sich bestimmte Fragestellungen ausschließlich auf der Grundlage bereits erhobener Daten beantworten lassen. Insbesondere wenn Veränderungen im Zeitverlauf (bspw. der Wandel von Einstellungen oder Verhaltensweisen in bestimmten Lebensbereichen) untersucht werden sollen, besteht insofern die Notwendigkeit auf vorhandene Daten zurückzugreifen, als dass man solche Daten nicht retrospektiv erheben kann. Darüber hinaus sind sozialwissenschaftliche Forschungsdaten, insbesondere solche, die mittels Befragungen oder durch beobachtende Verfahren entstehen, nicht reproduzierbar. Da sie gesellschaftliche Verhältnisse zu einem gegebenen Zeitpunkt beschreiben, sind sie somit auch für die historische Forschung von grundsätzlicher Bedeutung. Per Definitionem handelt es sich bei der Nutzung von Daten der amtlichen Statistik und von prozessproduzierten Daten für wissenschaftliche Forschungszwecke immer um Nachnutzungen, da diese ursprünglich nicht zu dem Zweck der Forschung erhoben wurden.

Metadaten

Metadaten spielen eine entscheidende Rolle für die Nutzbarkeit sozialwissenschaftlicher Forschungsdaten. Insbesondere die Nutzung durch Dritte, die nicht unmittelbar am Entstehungsprozess der Daten beteiligt waren, ist ohne umfassende Dokumentation kaum möglich. Hinzu kommt, dass sozialwissenschaftliche Daten häufig auch viele Jahre oder gar Jahrzehnte nach ihrer Entstehung noch eine hohe Relevanz für die Forschung haben können (bspw. zur Untersuchung sozialen Wandels). Die sozialwissenschaftlichen Datenarchive stellen ihre Dokumentationen lange Zeit in der Form von sogenannten Codebüchern bereit. Diese

enthielten neben Informationen zur Entstehung der Daten (z.B. beteiligte Primärforscher, Erhebungszeitraum, Beschreibung der Stichprobenziehung und der verwendeten Erhebungsmethoden) auch eine detaillierte Beschreibung des Datensatzes (z.B. Position von Variablen im Datensatz, Messniveaus, Bedeutung der numerischen Codes) und der verwendeten Messinstrumente (z.B. Frage- und Antworttexte der Fragebogen, Filterführungen, Intervieweranweisungen). Mit dem Ziel eine qualitativ höherwertige und insbesondere standardisiertere und strukturiertere Form der Dokumentation zu erreichen, formierte sich 1995 eine internationale Gruppe von Datenarchiven und -produzenten zur Data Documentation Initiative (DDI) und legte ein Jahr später die erste Version der DDI-Metadatenpezifikation vor.²⁸ Die ersten Versionen der Spezifikation orientierten sich noch sehr stark am klassischen Codebuch der Datenarchive. Mit der aktuellen Version 3 (bzw. 3.1) wurde die Spezifikation mit ihrer Ausrichtung auf den kompletten Forschungsdatenzyklus – von Design und Planung über Datenerhebung, -aufbereitung und -analyse bis hin zur Archivierung – und einem auf XML-Schemata basierenden modularen und erweiterbaren Aufbau fundamental verändert.

Für die Domäne der statistischen Aggregatdaten ist neben DDI insbesondere SDMX (Statistical Data and Metadata Exchange)²⁹ von Bedeutung und in Verwendung. Die Entwicklung des Standards basiert auf der Initiative von sieben inter- bzw. supranationalen Institutionen (u. a. OECD, Weltbank und EZB), die 2005 in der Veröffentlichung der ersten Version des Standards mündete.

Darüber hinaus ist als domänenübergreifender Metadatenstandard noch das kürzlich in der Version 2.1 veröffentlichte Metadatenschema von DataCite zu nennen. Dieses umfasst einen Bestand an deskriptiven Metadaten, der in erster Linie eine konsistente Identifizierung und Zitation von Forschungsdaten unterstützen soll. Mit dem Ausbau der im DataCite-Verbund angesiedelten Registrierungsagentur *datacite.org* und der vermehrten Vergabe von DOIs für sozialwissenschaftliche Daten wird dieser Metadatenstandard für die Referenzierung und Recherche von sozialwissenschaftliche Forschungsdaten an Bedeutung gewinnen.

28 Vgl. DDI (2011).

29 Vgl. SDMX (2011).

4.4 Interne Organisation

Auch in den Sozialwissenschaften gehört Datenmanagement nicht zum Ausbildungsprogramm der Fachwissenschaftler, weshalb in universitären Forschungsprojekten in der Regel keine systematische Langfristsicherung erfolgt. Durch die bereits lange Tradition der sozialwissenschaftlichen Datenarchivierung gibt es jedoch in den wenigen vorhandenen speziellen Institutionen seit langem bewährte Arbeitspraxen, viel Erfahrung mit Migrationsprozessen und etablierte Qualitätskontrollen z.B. der Metadatenaufbereitung. Diese wurden etwa durch die Kooperation im CESSDA-Verbund in internationaler Abstimmung weiterentwickelt und der Community zur Verfügung gestellt.³⁰ Dennoch hat bisher die selbst auf internationaler Ebene absolut gesehen geringe Mitarbeiterzahl des spezialisierten Arbeitsfeldes eine durchaus erwünschte Professionalisierung gebremst und auch die Formalisierung der Arbeitsprotokolle ist nicht überall befriedigend entwickelt. Die seit einiger Zeit neu entstehenden Normen und Standards zur Langzeitarchivierung (wie bspw. OAIS³¹, „Audit and certification of trustworthy digital repositories“³² oder auch der DIN-Norm-Entwurf „Kriterien für vertrauenswürdige digitale Langzeitarchive“³³) werden daher aufmerksam beobachtet und in ersten Schritten implementiert.

Schon die interne Standardisierung und Qualitätssicherung von Prozessen erfordert eine entsprechende Qualifizierung des Personals. Dies gilt umso mehr, als nach langjähriger Erfahrung der Autoren der Arbeitsmarkt bislang kein Personal mit der geeigneten Misch-Qualifikation aus technischer, archivarischer und sozialwissenschaftlicher Expertise bietet, sondern diese durch mehrjähriges *training on the job* erst aufgebaut werden muss. Dies wiederum nehmen die Mitarbeiterinnen und Mitarbeiter vor allem dann mit der nötigen Motivation wahr, wenn ihnen in der Institution eine langfristige Perspektive geboten werden kann, da das zu erwerbende Spezialwissen auf dem externen Arbeitsmarkt nur schwer einsetzbar ist. Im GESIS Datenarchiv für Sozialwissenschaften als dem ältesten spezialisierten Dienstleister für wissenschaftlichen Datenservice in den Sozialwissenschaften ist

30 Vgl. CESSDA (2011b).

31 Vgl. ISO 14721:2003.

32 Vgl. ISO/DIS 16363.

33 Vgl. DIN 31644.

eine klare Mehrheit der Beschäftigungsverhältnisse unbefristet.³⁴ Diese aus den genannten Gründen notwendige hohe Mitarbeiterbindung ist aber nur möglich, wenn der Arbeitgeber selbst langfristig finanziert ist.

Das GESIS Datenarchiv, welches als Teil vom GESIS-Leibniz-Institut für die Sozialwissenschaften nach den für die Leibniz-Gemeinschaft typischen Modalitäten in einer langfristigen Bund-Länder-Finanzierung steht, erhebt von datengebenden Primärforschern keinerlei Gebühren für die Archivierungsleistung. Die inzwischen weit überwiegend genutzte Datenbereitstellung über Online-Systeme an Datennutzer ist ebenfalls gebührenfrei. Allein für manuelle Datensatzzusammenstellungen und Postversand werden geringfügige Handhabungsgebühren erhoben, die keinen relevanten Finanzierungsbeitrag darstellen.

Seit einigen Jahren (s.o.) sind als bisher meist mittelfristig finanzierte Projekte Forschungsdatenzentren hinzugekommen, die überwiegend eine Anschubfinanzierung durch das BMBF erhalten haben. Nach Auslaufen dieser Anschubfinanzierung ergeben sich unterschiedliche Finanzierungsmodelle. So werden die Forschungsdatenzentren der Länder und des Statistischen Bundesamtes in die Trägerschaft ihrer jeweiligen Mutterinstitutionen übernommen, müssen seit 2011 aber einen Teil ihrer Kosten durch Gebühren auf den Datenservice decken. Diese Gebühren sind grob aufwandsabhängig und können nach den Maßstäben der Sozialwissenschaften beträchtlich sein.³⁵ Andere Forschungsdatenzentren, wie jene der GESIS, das FDZ SOEP am DIW Berlin, das FDZ PsychData am ZPID Trier, sind erst in jüngerer Zeit als integraler Teil ihrer Mutterinstitute (hier jeweils Institute der Leibniz-Gemeinschaft) etabliert worden und unterliegen somit denselben Finanzierungsbedingungen wie diese Mutterinstitute. Mit dem FDZ pairfam ist zudem jüngst ein Forschungsdatenzentrum im Rahmen eines DFG-Langfristvorhabens entstanden, womit eine (nach Ablauf erneuerbare) Befristung auf einen Förderzeitraum von maximal zwölf Jahren verbunden ist.

34 Bei den ca. 15 technischen Sachbearbeiterstellen des GESIS Datenarchivs gelten ohnehin die üblichen Entfristungsregeln des Arbeitsrechts, bei den gut 30 wissenschaftlichen Mitarbeitern (u.a. einschließlich Mitarbeitern von drei Forschungsdatenzentren bei GESIS) kommt das Wissenschaftszeitvertragsgesetz zur Anwendung.

35 Vgl. Statistische Ämter des Bundes und der Länder – Forschungsdatenzentren (2011a); RatSWD (2010c).

Für Kooperationsbeziehungen zu Zwecken der Langzeitarchivierung gibt es keine vorgegebenen Modelle. Einzelne Forschungsdatenzentren haben bereits Vereinbarungen über eine Langzeitsicherung ihrer Daten mit dem GESIS Datenarchiv, mit weiteren FDZ gibt es Gespräche. Für das breite Feld der universitären Forscher ist weiterhin das GESIS Datenarchiv eine zentrale Institution zur Langfristsicherung, die vor allem für Umfragedaten stark genutzt wird, indem ca. 130 neue Studien pro Jahr deponiert werden. Auch größere Gruppen wie DFG-geförderte Sonderforschungsbereiche haben die erzeugten Datenbestände dem GESIS Datenarchiv zur Langfristsicherung und weiteren Bereitstellung überantwortet oder sind im Begriff dies zu tun.

4.5 Perspektiven und Visionen

Zwar haben die Sozialwissenschaften mit den oben dargestellten Institutionen und dem robusten Angebot öffentlicher Aggregatdaten durch statistische Ämter usw. bereits einen vergleichsweise guten Reifegrad der Infrastrukturangebote erreicht. Es kann dennoch leider nicht behauptet werden, dass data sharing für sozialwissenschaftliche Forscher aus allen Disziplinen bereits eine Selbstverständlichkeit sei. Selbst für Umfragedaten ist die Abdeckung nur partiell, und z.B. für qualitative Daten sind entsprechende Strukturen erst in den Kinderschuhen. In den Wirtschaftswissenschaften scheinen die beim Rat für Sozial- und Wirtschaftsdaten akkreditierten Forschungsdatenzentren eher Ausnahmecharakter zu haben, insofern als für das breite Feld akademischer Einzelprojekte Datensicherungsangebote erst in der Planungsphase sind.³⁶

Dies hat sich auch durch die DFG-Empfehlungen zur guten wissenschaftlichen Praxis nicht wesentlich geändert. Die dort ausgesprochene Empfehlung 7 zur Aufbewahrung von Primärdaten stellt mit der Maßgabe einer Aufbewahrungsfrist von nur zehn Jahren innerhalb der datenerzeugenden Institution allein eine gewisse Vorkehrung für gezielte Replikationsbemühungen veröffentlichter Forschungsergebnisse dar.³⁷ Im

36 Persönliche Kommunikation von Markus Quandt mit Olaf Siegert, Leibniz-Informationszentrum Wirtschaft (ZBW) über dort laufende Projektplanungen.

37 Vgl. DFG (1998), S. 12.

Sinne der Förderung von data sharing muss man hingegen eher Sorge haben, dass sich die Empfehlung nachteilig auswirkt, denn sie könnte bei Datenproduzenten das Missverständnis erzeugen, dass eine befristete lokale Datensicherung ohne Einhaltung irgendwelcher Standards bereits hinreichend sei, um aktuellen wissenschaftlichen Normen zu genügen. Vorbildlich ist in dieser Hinsicht das britische Economic and Social Research Council, welches bei von ihm geförderten Projekten fordert, dass neu erzeugte Daten den relevanten Datenservice-Instituten kurz nach Projektende zur Archivierung angeboten werden.³⁸ Die Nichteinhaltung dieser Auflage ist mit finanziellen Sanktionen belegt.

Einen wesentlichen Anreiz zur Bereitstellung selbsterhobener Forschungsdaten dürften in Zukunft aber auch Systeme setzen, die veröffentlichte Forschungsdaten dauerhaft zitierbar machen und somit den Urhebern dieser Daten einen Reputationsgewinn aus solchen Zitationen ermöglichen. Da sich mehrere Forschungsdatenzentren der von GESIS und ZBW gemeinsam betriebenen Registrierungsagentur da|ra bedienen werden,³⁹ sind die Chancen auf eine weitere Durchsetzung dieses Lösungsansatzes in den deutschen Sozialwissenschaften gut.

Mit einer zentralen Datenregistrierung wird zudem auch ein Problem wesentlich gemildert, welches durch die Eröffnung vieler Zugänge zu neuen Daten (wie es mit dem Erfolg der Forschungsdatenzentren geschieht) erst entstanden ist, nämlich das der wachsenden Schwierigkeit für Nutzer, die Übersicht über die unterschiedlichen Datenquellen zu behalten.

Ein weiteres Problem wird in den Sozialwissenschaften derzeit noch vor allem auf der internationalen Ebene sichtbar. Vor den Zeiten grenzüberschreitender Online-Zugriffsmöglichkeiten wurden inhaltsgleiche Datensätze bewusst parallel in den Beständen verschiedener nationaler Archive geführt, um sie für die jeweils nationalen Nutzerkreise schnell im Zugriff zu haben. Weil inzwischen eine digitale Bereitstellung auch über nationale Grenzen hinweg sehr einfach ist, überwiegen nun die Nachteile derartiger Parallelbestände. Diese Nachteile bestehen darin, dass zahlreiche sozialwissenschaftliche Datensätze – oft gerade die in der Forschung

38 Vgl. ESRC (2010).

39 Stand: Frühjahr 2011.

am stärksten nachgefragten – häufigen Aktualisierungen und Versionsänderungen unterliegen und daher die Identität zweier digitaler Datensatz-Objekte an unterschiedlichen Orten nur mit erheblichem Aufwand kontinuierlich sichergestellt werden könnte. Die einfachste Lösung dazu ist, eindeutige Verantwortlichkeiten für die Versions- und Nachweispflege („*data curation*“) jedes relevanten Datensatzes an einzelne Archive zuzuweisen. Solche Vereinbarungen können aber derzeit nur bilateral geschlossen werden. Hierzu werden multilaterale internationale Vereinbarungen und verbindliche Konfliktlösungsmechanismen erforderlich werden.

Grundlegender dürfte jedoch sein, das Idealbild eines Datenlebenszyklus mit bruchlosen Übergängen zwischen den Phasen des Untersuchungsdesigns, der Datenerhebung, der Aufbereitung, Archivierung und Bereitstellung zur Sekundäranalyse in die Praxis umzusetzen. Dazu ist nicht nur auf Seiten der Datenproduzenten ein Fortgang des Bewusstseinswandels erforderlich, sondern es muss auch von den Dienstleistern in den Dateninfrastrukturen stark darauf geachtet werden, den jeweiligen Fach-Communities entgegen zu kommen. So ist z.B. die Erfassung von hinreichenden Metadaten in der Regel nur mit direkter Hilfe der Primärforscher möglich. Zwar können diese partiell von der frühzeitigen Anwendung von Metadatenstandards profitieren, etwa beim Design ihrer Forschungsinstrumente, falls entsprechende standardkompatible Softwaretools wie etwa Fragebogeneditoren mit DDI-Unterstützung verfügbar sind. Weil aber auf breiter Front von den Primärforschern eine Einarbeitung in die Verfahren des ingest oder in Verschlagwortungsthesauri kaum erwartet werden kann, ist in der Regel eine gezielte Unterstützung durch Datenmanagementspezialisten frühzeitig im Forschungsprojekt notwendig.

Um in Zusammenarbeit mit der großen Anzahl von Datenproduzenten operable Verfahren und Standards des Datenmanagements entwickeln und implementieren zu können wie auch zur Ausbildung entsprechenden Fachpersonals wären einige wenige Kompetenzzentren für Forschungsdatenmanagement in den Sozialwissenschaften erforderlich. Diese Zentren müssen untereinander stark kooperieren. Auch bei Methoden der reinen Langfristsicherung wird z.B. im Zuge von Format-Migrationsprozessen, die wegen der Formatvielfalt sozialwissenschaftlicher Datenanalysesysteme erforderlich sind, weiterhin fachwissenschaftliche

Expertise unabdingbar bleiben. Allein die rein IT-technische Dienstleistung der *bitstream preservation* kann daher an Cloud-Dienste, Großrechenzentren oder ähnliches verlagert werden, soweit dies kosteneffizient ist und Vorgaben des Datenschutzrechtes dies nicht verbieten.